ATTACHMENT II

PROTEINS: Structure, Function, and Genetics 23:301–317 (1995)

A Critical Assessment of Comparative Molecular Modeling of Tertiary Structures of Proteins*

Steven Mosimann, Ron Meleshko, and Michael N.G. James
Medical Research Council of Canada, Group in Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada

ABSTRACT In spite of the tremendous increase in the rate at which protein structures are being determined, there is still an enormous gap between the numbers of known DNA-derived sequences and the numbers of three-dimensional structures. In order to shed light on the biological functions of the molecules, researchers often resort to comparative molecular modeling. Earlier work has shown that when the sequence alignment is in error, then the comparative model is guaranteed to be wrong. In addition, loops, the sites of insertions and deletions in families of homologous proteins, are exceedingly difficult to model. Thus, many of the current problems in comparative molecular modeling are minor versions of the global protein folding problem. In order to assess objectively the current state of comparative molecular modeling, 13 groups submitted blind predictions of seven different proteins of undisclosed tertiary structure. This assessment shows that where sequence identity between the target and the template structure is high (> 70%), comparative molecular modeling is highly successful. On the other hand, automated modeling techniques and sophisticated energy minimization methods fail to improve upon the starting structures when the sequence identity is low (~30%). Based on these results it appears that insertions and deletions are still major problems. Successfully deducing the correct sequence alignment when the local similarity is low is still difficult. We suggest some minimal testing of submitted coordinates that should be required of authors before papers on comparative molecular modeling are accepted for publication in journals. © 1995 Wiley-Liss, Inc.

Key words: molecular model, comparative model, homology model, structure prediction, calculated structure

INTRODUCTION

Once a protein's sequence has been determined and it has been found to be a new member of a structurally characterized protein family, it is relatively straightforward to build a molecular model of the protein using a set of simple guidelines.[1,2] Presently, there are several commercial and public domain computer programs that have been developed for modeling; these programs remove much of the tedium from the process. There are numerous reasons for constructing comparative molecular models of proteins. The molecular model may explain the structural basis of existing experimental results and can provide one with structural information on which further experiments can be planned, executed, and evaluated. Site-specific mutations of the gene coding for the specific protein can provide important data regarding the protein's function. Perhaps, some of the most revealing experiments are those designed to predict and to probe the molecular reasons for an enzyme's specificity.[3] On a more practical note, the molecular model can sometimes be used successfully to determine phases for a crystal structure determination using the method of molecular replacement.[4] The more spectacular uses, however, are typified by the recent successful application of comparative molecular modeling for identifying new classes of lead compounds in antimalarial drug development.[5]

An example of the successful prediction of an enzyme's specificity from comparative molecular modeling is that for granzyme B (CCP1), a serine proteinase from cytotoxic T lymphocytes.[3] A molecular model of CCP1 (48% identical to rat mast cell proteinase II) showed that an arginine at position 226 would occupy the S_1 specificity pocket, thereby suggesting a P_1 specificity for an aspartate or glutamate residue. Subsequent synthesis and testing of a series of substrates differing in the nature of the P1 residue confirmed the aspartate specificity of CCP1.[6] The P_1 specificity of CCP1 has recently been altered by site-specific mutagenesis of the residue at

*This assessment does not indicate that any one particular modeling group or modeling technique is superior to any other. We do not believe that comparative molecular models can be ranked using a single or even several numeric indicators. As such, claims that particular modeling techniques are superior based upon the results herein are not justifiable, in our opinion.

Received March 30, 1995; revision accepted June 20, 1995.
Address reprint requests to Michael N.G. James, Medical Research Council of Canada, Group in Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada.